

# Leveraging Social Network Analysis for Characterizing Cohesion of Human-Managed Animals

Dixon Vimalajeewa, Sasitharan Balasubramaniam, Bernadette O'Brien, Chamil Kulatunga, Donagh P. Berry  
 dvimalajeewa@tssg.org, sasib@tssg.org, Bernadette.O'Brien@teagasc.ie, chamilkul@gmail.com,  
 Donagh.Berry@teagasc.ie

**Abstract**—Social network analysis is a technique to study behavioral dynamics within a social group. In social network analysis, it is an open question as to whether it is possible to characterize animal-level behaviors by using group-level information. Also, it was believed that the combined use of social network analysis would provide a more comprehensive understanding about social dynamics. In light of these two factors, here we explain an approach to evaluate animal importance to a group by considering the variability in group-level structural information, which is computed by joining the animal- and group-level social network analysis measures node centrality and network entropy, respectively. Moreover, two other metrics, animal social interaction range and nearest-neighbor frequency matrix, which represent social affiliation of each animal within the group, are computed to help address the general challenges in graph-based social network analysis, and thereby, improve the precision of animal importance measures. Finally, we derive the joint distribution of animal importance of the group in detecting atypical social behaviors. The approach is tested using tracking data of dairy cows. The reliability of the derived animal importance was superior to the already existing animal importance measures. To illustrate the usability of the animal importance metric, a simulation study was conducted to identify sick and estrus animals in a group. The social affiliation of sick cows was less when compared to healthy cows. Also, their individual distributions of animal importance were shifted towards the left of the mean of the animal importance distributions of healthy cows. Consequently, the joint distribution of animal importance of the group exhibited a bi-modal distribution with a left tailored shape. The behavior of cows in estrus was opposite to that of sick cows. Moreover, with increasing number of sick and estrus cows in the group, respectively the group entropy decreased with larger variance and slightly increased with less variance. Therefore, the entropy-based animal importance metric has superior performances when evaluating animal importance to the group compared to the existing metrics. It can be used

for generating alerts for the early-detection of atypical social behaviors associated with, for instance, animal health, veterinary, and welfare.

**Keywords**—social network analysis, animal importance, social interaction range, social affiliation.

## I. INTRODUCTION

Farm animals are usually gregarious animals and their social interactions could greatly be capitalized on improving farm management operations such as improving animal well-being [1]. Evidence that supports the significance of their social relationships is considerably under-utilized [2]. The growing interest in social network analysis (SNA), which is a popular method to study the behaviours social groups, confirms that it can offer great advantages by exploring individual social affiliations [3]. Also, recent advancement in animal monitoring systems which enable the collection of vast amount of data, such as wireless sensor networks, facilitates large-scale SNA in broader perspectives [4]. Consequently, this necessitates the investigation of novel strategies for SNA to transform animal social relationships into useful metrics, which can subsequently be used to support better farm management practices.

Modern-day animal production systems routinely exploit state-of-the-art technologies to support decision making by generating valuable insights from data collected by sensors. Most pervasive sensor technology is the Global Positioning System (GPS), which determines geo-location [4], and therefore the movements of individual instrumented animals. Other energy-efficient techniques such as Wi-Fi, bluetooth, and LoRA wireless signal strength based positioning technologies are also becoming increasingly popular [5], [6]. Therefore, opportunities exist to define novel and informative insights (e.g., phenotypes) from such geo-location data. However, the necessary descriptive metrics need to be defined [7] from which deviations can be identified and subsequent alerts generated. Such metrics should be informative to characterize not only the animal itself, but also the group or sub-group level behaviors. Once the behaviors of an individual animal are characterized relative to the group, any deviations from the norm can be used as early and real-time alerts for producers to identify, for instance, atypical behaviors of animals that may require particular attention.

---

D. Vimalajeewa is with the Telecommunications Software and Systems Group, Waterford Institute of Technology, Waterford, Ireland.

S. Balasubramaniam is with Telecommunications Software and Systems Group, Waterford Institute of Technology, Waterford, Ireland and Faculty of Information Technology and Communication Sciences, Tampere University, Finland.

C. Kulatunga is with the University College Dublin, Ireland.

D. P. Berry and B. O'Brien are with the Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland.

This work was supported by the Science Foundation Ireland (SFI) projects PrecisionDairy (ID: 13/1A/1977) and VistaMilk (ID: 16/RC/3835), and the Horizon 2020 GenTORE project. E-mail: dvimalajeewa@tssg.org

Animals living in groups are generally social animals and have complex relationships and social behaviors [8]. For instance, cows are social animals [9] and prefer to stay in groups, synchronizing their activities mostly based on neighboring (or dominant) animals. Although external stimuli can perturb normal group behaviors of such animals, there is a tendency for the animals to revert to normal status as soon as possible [10]. This is because of animal cohesion, which is a phenomenon that describes how strongly animals are connected to each other and is used to understand behavioral dynamics based on their heterogeneous social relationships in the group [9]. Changing such dynamics is mostly influenced by a small set of group members generally identified as the most important (or highly cohesive/interactive) animals to the group [11]. Hence, understanding individual's importance to the group would be useful for exploring different behaviors of groups as well as individuals such as group synchronization, spreading dynamics, and cascading reactions [12].

SNA provides a comprehensive understanding of social dynamics among animals and also facilitates quantifying them [12], [13]. Graph theoretical concepts are commonly used for SNA. A group is depicted as a graph in which nodes represent individual animals, and an edge represents the interaction between two animals (Figure 2). This model is therefore capable of representing the complex structure of interactions among individuals, also called a sociogram [12]. The heterogeneous nature of interactions is then used to determine the importance of each animal. The importance of an animal to the group is recognized as its involvement in forming a complex network structure, i.e., to what extent that animal is connected to the others in the network, and quantified as the centrality. Importance is generally derived by exploiting an individual's local level structural information (only within the social interaction range) such as degree (i.e., the number of direct edges on each node) and closeness (i.e., the shortest path length between a particular edge with all other edges) [12]. Therefore, the temporal evolution of centrality of group living animals is considered as one of the relevant characteristics for defining metrics which can subsequently be used to describe different behavioral patterns in SNA [14].

Different ideas have been highlighted in order to assist in developing novel metrics for SNA. Among those ideas, [12] stated that the combined use of local, intermediate, and global measures would provide a more comprehensive understanding of the network dynamics in a broader view; based on the level of information used to quantify different social characteristics where graph measures are divided into three categories as local, intermediate, and global (or graph). Besides, [15] unveiled that using a broader range of information in quantifying network properties would provide superior performances than being limited to the domain that they have been defined for. So, it is an open question as to whether it is possible to derive local (animal)-level measures based on the global (group)-level structural variability of social dynamics [16]. Therefore, the present study is an attempt to apply all these concepts together in order to derive a metric to evaluate animal importance (or sociality) to a group ( $AIM$ ) based on the dairy cattle mobility observations. An additional aim is to explore the possible

opportunities, which can be benefited by using the derived  $AIM$ , in relation to the dairy cattle behaviors as a use case.

There are different challenges, however which cannot be disregarded when deriving novel measures, such as spatial and temporal limitations, defining the network and sampling animals into it, and validity and robustness of derived network measures [12]. So, prior to conducting a SNA, careful attention should be given to those factors and the necessary actions should be in place to mitigate against them. Although graph-based SNA generally assumes that the interactions included in the graph are relatively stable over time, that is not always true because animal relationships are highly dynamic and so will the network topology. Also, sampling relevant animals into the network including their significant interactions, is vital as if there are too many animals, the network might split into unconnected subgroups. This may cause complications in applying graph theoretic measures such as diameter, which could consequently lead to misleading information. Practically, animal interactions depend on various factors such as age, gender, and health, so that selecting only the significant interactions is crucial [11]. Disregarding such interactions could negatively impact on the representativeness of the network and so will the robustness of the derived measures. Consequently, not only would the reliability of the estimated network dynamics become invalid, but it would also contribute to misleading outcomes about social behavioral dynamics.

Therefore, the main contribution of this study is to derive an approach to evaluate  $AIM$  by combining local- and global-level measures, node centrality and network entropy, addressing the SNA challenges mentioned above. While network entropy depicts the amount of information encoded within a network and is used to compute the structural complexity at the group level [17], the node centrality quantifies node importance based on the extent to what a node is surrounded by other nodes [18]. In this approach, the influence of an animal on changing the network entropy is considered as its  $AIM$ . The idea is to taking into account the variability of graph-level structural information in evaluating node-level properties. Thus, this approach facilitates expanding the range of information used in quantifying  $AIM$ . Additionally, this process is backed by two other metrics, animal interaction range and nearest-neighbor (NN) frequency matrix. The animal interaction range, which stands for the optimum range where an animal can make strong interactions, helps to sample significant interactions into the network graph. Therefore, it helps improving the representativeness of the network graph and enhancing the validity of  $AIM$ . The NN frequency matrix demonstrates the social affiliation of each animal with others in the group as a frequency value counted over the time. As this matrix represents the preferential members of every animal, it is easy to recognize animals which have strong interactions in the group. We use these approaches in demonstrating detection of sick and estrus (sexually active) cows in dairy herds as a use case of our approach. The variation of the normal probability density function (PDF) of  $AIM$  at individual animal-level as well as the Gaussian mixture model of  $AIM$  at group-level were explored to identify sick and estrus animals from the normal animals in the herd.

The remainder of the paper is organized as follows. While Section II discusses related works, Section III explains the theoretical steps of deriving *AIM*. In Section IV, the approach described in Section III is applied to a real dairy cattle mobility dataset. Section V demonstrates the applicability *AIM* in detecting sick and estrus cows from a herd based on the simulated mobility data, including some directions to continue this study further, and Section VI concludes the paper.

## II. RELATED WORKS

The use of SNA has gained considerable attention in a wider range of fields such as sociology, business, and ecology [19], [12]. Exploring social inter-relationships, quantifying disease transmission, and building models to explain dynamics in network topology, are some of the highlighted applications. The study [12] emphasized that graph-based SNA is a promising tool for exploring such applications. Identifying and forming a network, developing methods for characterizing social behaviors, and exploring dynamic variability in social inter-relationships are some of the key areas that animal science has been benefitted from SNA. Also, SNA has increasingly been used for improving the efficiency of human-managed animal farms.

Forming the network is one of the most important and challenging tasks in SNA. Because not only are social interactions heterogeneous, but also are the factors which influence making interactions. Therefore, sampling animals in a network and defining their interactions must be done carefully as they contribute to improving the representativeness of the animal group and so the reliability of the derived social behaviors. One of the simplest ways of measuring social interaction is the use of NN identity. [20] used NN identity data for assessing interactions based on a clustering technique in which a local group cluster matrix was developed in order to identify the most interactive nodes in a network. Evaluating the strength of the interactions among nodes is vital for conducting a comprehensive SNA. While the study [21] proposed weighted degree and strength centrality measures, taking into account the weight (strength) distribution of interactions, [11] discussed different mathematical approaches which can be used in computing the strength of the interactions in constructing the network.

The greater opportunities for monitoring social behaviors in a real-time manner has necessitated the development of novel SNA approaches through re-examining already existing techniques. As a result, various attempts have been made in deriving novel metrics in different applications. For instance, [15] developed a novel measure called Laplacian centrality to compute node importance and proved that it has greater performances than the standard centrality measure based on terrorist network analysis. Following the work in [15], the study [16] proved that using network-level information in quantifying node-level attributes can significantly improve the accuracy of selecting the top- $k$ -most important nodes compared to the existing measures. These studies were, however, based on the static networks, but measures, which enable capturing time variant features, were highly demanded in SNA.

Therefore, most research on SNA has focused on the exploring dynamic properties of social networks. For instance, while the study [18] derived a novel metric, dynamic centrality by exploring the limitations accounted in static network graph-based SNA, the time-ordered-graph method explained in [22] converts a dynamic network into a static networks enabling the application of static SNA measures.

With the advancement of precision agriculture, applications of SNA for human-managed animals have drawn a considerable attention. For instance, significant improvements in farm management such as individual animal fitness, controlling disease transmission, and welfare could be achieved by exploring the social behavior of animals over a long period of time emphasizing automatic location measurements, NNs, and NN distance in SNA with farm animals [23], [24], [25]. The studies [1] and [2] emphasized the importance of using SNA in particular to dairy cattle management by exploring different characteristics such as community structure, social differentiation, stress, and productivity. However, [2] highlighted that SNA concepts have widely been used for characterizing the different social behaviors of wild animals, but not much for human-managed animals. Therefore, there are opportunities that SNA can be used in intensifying farm operations though, they are not yet fully realized by the wider research community.

Our hypothesis is that SNA can effectively be capitalized in understanding and quantifying various factors such as husbandry practices, health issues, feeding, stress, and survival, thus making a significant impact on the stability of farm production systems and decision support tools. Moreover, the scale in which SNA can be applied is continuously increasing within the modern data monitoring tools. Hence, the necessity of deriving novel SNA approaches for applying to large-scale applications is also emphasized.

## III. THEORETICAL BACKGROUND

This section explains the process of deriving the entropy-based *AIM*. Initially, the procedure of computing the nearest-neighbor frequency matrix is discussed. This is followed by an experiment of the topological distance based animal interaction range and we then discuss how to use network entropy to validate the interaction range. These two measures are then used in the process of computing animal importance to the group (i.e., *AIM*). Figures 1 and 2 together give a graphical overview of the process in four steps. Then, in order to represent the variation of importance of individual- and group-level in a distributional sense, the Probability Density Function (PDF) and Gaussian mixture model of *AIM* are discussed. Finally, to test the validity of the derived measures, the system used to collect data and how the experiment was conducted, is explained.

### A. Nearest-Neighbor Graphs

The spatial variability of nearest-neighbors (NNs) around each animal is commonly used to identify the most interactive individuals [20]. Therefore, the NN approach was used as the foundation to understand the evolving cohesion (or social

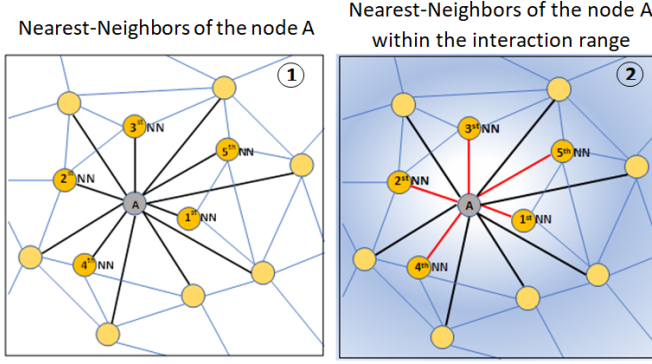


Fig. 1: Nearest-neighbors and interaction range.

affiliation) with respect to each animal's daily activities and its interactions with other animals in the group.

We consider the mobility of a group of  $N$  animals over a period of time  $T$ . Then we quantify the variation of animal relationships in every time window  $t (\leq T)$ :  $T_w = T/t$  number of time windows. Information on NNs for each animal in the present study was determined based on the geographical distance between every pair of individuals in every  $t$  minutes time window (Figure 1(1)). The frequency of each animal being the  $k^{th}$  NN ( $0 < k < N, k$  is the order of NN) to every other animal in the herd within the entire time period was formulated as a frequency matrix ( $A_{N \times N}$ ). The  $k^{th}$ -NN algorithm was used to identify the NNs of each animal in the group based on the inter-animal geographical distances. The "haversine" formula [26] was used to compute the shortest-distance between each of the locations of two animals. The matrix  $A$  was therefore defined as:  $A = [f_{i,j}]$ , where  $f_{i,j}$  means the frequency of the  $j^{th}$  animal being a NN to the  $i^{th}$  animal over the period  $T$ . The matrix  $A$ , however is not a symmetrical matrix, since interactions are not always symmetric. For instance, suppose the two animals  $i$  and  $j$ , and  $j$  is the  $1^{st}$  NN of  $i$ ; then, the  $1^{st}$  NN of  $j$  would not be  $i$  whenever there is another animal closer to  $j$  than  $i$ . Hence, the interaction between  $i$  and  $j$  would not be symmetrical. The  $i^{th}$  row of  $A$  represents the spatial variation in frequencies of the other  $N - 1$  animals being a NN to  $i$ .

In practice, it cannot be considered that each animal interacts with all other animals in the group in a particular time window because the strength of interactions would be very weak with distant animals (i.e., higher order NNs). Therefore, on average, the optimal number of interactions that every animal can have (i.e., the number of strong interactions) is essential in studying the intensity of interaction frequency of each animal within the herd and also constructing social network graph (Figure 1(2)). The interaction range metric derived below explains how to select those interactions.

### B. Animal Interaction Range ( $k_r$ )

Sampling the most significant interactions in a network graph is one of the most crucial steps in graph-based SNA

because the quality of the network graphs affects the robustness of social characteristics derived from them. The social influence range is the region where an animal exhibits strong interactions, and it depends on the individual's sociality within the group. Since sociality depends on various factors such as gender, age, and health, the social influence range can vary from animal to animal [25]. So, an optimal measure to quantify social influence range is essential. Therefore, the main advantage of having an accurate social influence range is that it minimizes the loss of animal connectivity information while selecting interactions into a network graph. In the present study, the social influence range was defined as an animal interaction range (denoted as  $k_r$ ) and represents the optimal number of interactions with high weights. The interaction weight was the reciprocal of the distance between a pair of interacting animals.

Although many studies such as [27] and [3] commonly used the metric distance for deriving the interaction range, [10] proved that topological distance, i.e., interaction range with number of animals, is more robust than the metric distance. Therefore, the present study used the topological distance approach to derive the interaction range. The advantages of using this approach are the density fluctuations of animal aggregation, which can be well explained compared to metric distance, and the metric distance, which can be derived from the topological distance. Therefore, the interaction range was quantified based on the animal anisotropy factor and the network entropy. The network entropy-based  $k_r$  was used to confirm the result obtained from the anisotropy factor.

1) *Anisotropy Factor ( $\gamma$ ) based  $k_r$  Estimation:* Based on the spatial distribution of NNs and their orientation around each animal, the anisotropic structure of a moving group of animals varies with increasing order of NNs (i.e.,  $k$ ). The anisotropy value ( $\gamma$ ) represents the effect of interaction among animals, whatever the interaction is, and quantifies to what extent the spatial variation of the  $k^{th}$  NN (around a reference animal) is anisotropic. This is subsequently used to determine  $k_r$ , regardless of the distance between the animals. Initially, animal locations in the present study were converted into three-dimensional Cartesian coordinates ( $X, Y, Z$ ) and a set of NN vectors ( $u$ ), which were derived as:

$$u^{(i,k)} \in u, u^{(i,k)} = [(u_x^i, u_y^i, u_z^i)]_{(1 \times 3)},$$

where  $u_{i,k}$  is a unit vector directed towards the  $k^{th}$  order NN of the animal  $i$  (here  $i = 1, 2, \dots, N$ ). Also, the animal group center velocity vector ( $V$ ) was also calculated as:

$$V = [v_x, v_y, v_z]_{(1 \times 3)},$$

$$v_x = \frac{1}{N} \sum_{i=1}^N \frac{x_i}{t}, v_y = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{t}, v_z = \frac{1}{N} \sum_{i=1}^N \frac{z_i}{t},$$

where  $t$  is the width of the time window. Then normalized  $u$  and  $V$  ( $u_n$  and  $V_n$ ) were computed as:

$$u_n^{(i,k)} = \frac{u^{(i,k)}}{\sqrt{(u_x^i)^2 + (u_y^i)^2 + (u_z^i)^2}}, \quad V_n = \frac{V}{\sqrt{v_x^2 + v_y^2 + v_z^2}}$$



Secondly, in order to calculate the orientation of the  $k^{th}$  order NN vector of the  $i^{th}$  animal (i.e.,  $u_{i,k}$ ) with respect to the  $V_n$ , the projection matrix  $M_{i,k}$  of the  $i^{th}$  animal was calculated as:

$$M_i^k = [u_n^{(i,k)} u_a^i]_{(3 \times 3)}, \quad \text{for } i = 1, 2, 3, \dots, N,$$

where  $u_a^i \in u_n^{(i,k)}$  for  $a = \{x, y, z\}$ . Then  $M_{i,k}$  of all animals was averaged ( $M^k = \frac{1}{N} \sum_{i=1}^N M_i^k$ ) to derive a matrix, which projects along the average direction of the  $k^{th}$  order NNs of all animals in the herd. The  $M_k$  corresponds to the most relevant directions in the space and its eigenvalues represent the number of NN vectors that can be detected along the direction of the corresponding eigenvectors [11]. The anisotropy factor for the  $k^{th}$  order NNs,  $\gamma(k)$  was calculated at the third step by taking the square of the scalar dot-product between the normalized eigenvector ( $e_k$ ), which corresponds to the minimum eigenvalue of  $M_k$  and the normalized herd center velocity ( $V_n$ ) as:

$$\gamma(k) = (e^k \cdot V_n)^2 \quad (1)$$

The process was repeated varying the NN order, i.e.,  $k = 1, 2, \dots, N$ , in every time window and averaged over the  $T_{ws}$  to obtain the variation in anisotropy  $\gamma$  with respect to the spatial distribution of NNs (i.e., with increasing order of NN). The eigenvector corresponding to the smallest eigenvalue was used here since there is a reduced probability to detect an animal's NNs along the direction of this eigenvector. Finally, according to [10], for the isotropic case, the value of  $\gamma$  is 1/3 (i.e., no interaction) and the  $k$  value at which this occurs is defined as the value of interaction range (i.e., the value of  $k_r$ ). Thus, interactions between the reference animal with its NNs were deemed not significant after the  $k_r^{th}$  order NNs (i.e., isotropically distributed). In other words, the topological distance threshold is the  $k_r$  value, which is the number of individuals deemed to be around an animal. The metric distance is the distance to the  $k_r^{th}$  NN from the focal animal. More details on the anisotropy factor and its derivation are given in [10] and [11].

2) *Entropy-based  $k_r$  Estimation*: Variability in the daily group entropy, which is the average entropy (explained in the next section) over all the time windows in the present study, was computed as follows. Network graphs were generated by increasing the NN order (i.e.,  $k = 1, 2, \dots, N - 1$ ) around each animal. The increasing rate of herd entropy was examined over the time period  $T$ , and the  $k_r$  value was selected in units of animals up to where there was no significant incremental increase in herd entropy.

### C. Network Graph-based Animal Importance to the Group

Network graphs illustrate the structural connectivity of social groups and are more complicated due to the heterogeneous nature of social interactions (Figure 2(1)). Therefore, including only the most relevant interactions, for instance, interactions belong to the interaction range (i.e., within  $k_r$ ), would help to simplify the network complexity and would also enhance the quality of the structural information about animal connectivity acquired for characterizing different social behaviors, such

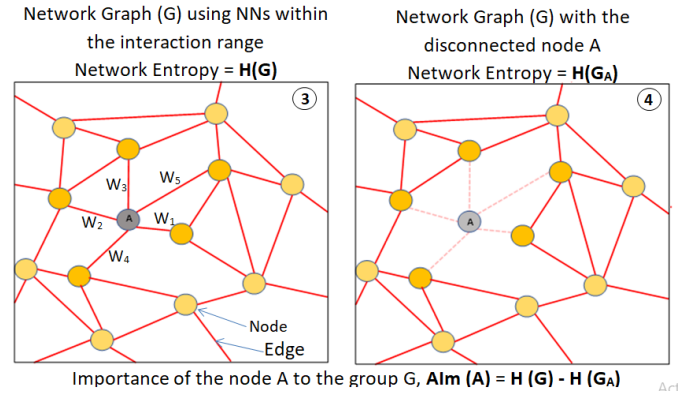


Fig. 2: Illustration of the technique of computing animal importance to a group,  $AIm$ .

as density, centrality, and communicability [12]. Such graph-theoretic measures can be applied directly to quantify these behaviors. Here, we follow an approach combining the network graph entropy and weighted degree centrality to evaluate animal importance to the group (denoted as  $AIm$ ). For instance, Figure 2(1) represents the network graph ( $G$ ) built by selecting the interactions, which belong to the interaction range of each node.

The variability in the number of interactions made by an animal depicts its influence to the group, because the removal of a highly interactive animal causes a substantial change in the structural group properties (e.g., group cohesion and connectivity). The animal importance to the group,  $AIm$  quantifies to what extent an animal influences the group. This can be measured as the group entropy variation caused by the removal of that animal from the group. The entropy ( $H$ ) characterizes the structural information based on a group level as well as on an individual animal level [16]. From a mathematical point of view, the entropy of a random variable  $X$ ,  $H(X)$  is the expected information encoded within  $X$  [28] and formulates as:

$$H(X) = E[I(X)]$$

where  $E$  is the expectation, and  $I(X)$  denotes the information contained in  $X$  and is computed as:  $I(X) = \frac{1}{P(X)}$ , where  $P(X)$  is the probability of  $X$ . Thus, according to the definition of statistical expectation,  $H(X)$  can be simplified into:

$$H(X) = -E[P(X)] = -P(X) \log P(X)$$

Following the entropy of a single random variable, the entropy of a system which consists of a set of  $n$  random variables can be computed as:

$$H(X) = - \sum_{i=1}^n E[P(x_i) \log(P(x_i))],$$

and this quantifies the expected information carried by the whole system.

To compute the entropy of the group (i.e.,  $H(G)$ ), a network graph was constructed considering only the interactions among individuals within the interactive range. Mobility of animals in the present study was considered as random variables and the amount of structural information attributable to each animal represents its influence to the herd. The direct weighted degree centrality ( $C_W$ ), which is a local (i.e., individual or node level) measure used in graph theory, was used to quantify an individual animal's influence on the herd. This measure reflects to what extent an animal is directly connected with other animals based on the weight of direct interactions [21]. There are two types of direct interactions associated with an animal: the interactions it initiates (i.e., outward) and the interactions it receives from others (i.e., inward). In the present study, the number of direct interactions was the sum of the both inward and outward interactions. Assuming an arbitrary time window  $t_i$ , suppose the  $i^{th}$  animal has  $v$  number of direct interactions with weights  $\{w_j\}_{j=1}^v$ ; the weight of an interaction was computed as the reciprocal of the distance between the pair of interacting animals. Then weighted centrality ( $C_W$ ) of the  $i^{th}$  animal,  $C_W(i)$  was computed as  $C_W(i) = \sum_{j=1}^v w_j$ , for  $i = 1, \dots, N$ , and then normalized ( $C'_W$ ) to compute the group entropy,  $H(G)$  as follows.  $C'_W$  satisfies all the conditions to be a probability distribution.

$$H(G) = \log \left( \sum_{i=1}^N C_W(i) \right) - \sum_{i=1}^N C'_W(i) \log(C_W(i)),$$

$$\text{where } C'_W(i) = \frac{C_W(i)}{\sum_{i=1}^N C_W(i)}$$

The importance of the  $i^{th}$  animal to the group,  $AI_m(i)$ , was computed as the change in group entropy caused by the removal of the  $i^{th}$  animal from the group, i.e.,

$$AI_m(i) = H(G) - H(G_i), \quad (2)$$

where  $G_i$  represents the group without the  $i^{th}$  animal. Also, Figure 2 graphically illustrates the computation of  $AI_m$  for the node  $A$ . The importance value of each of the  $N$  animals was computed for all time-windows and then averaged over the time period  $T$  to obtain the importance metric for that animal.

To compare the reliability of entropy-based  $AI_m$  with the other existing measures, the  $AI_m$  value of all animals was computed using the  $C_W$ , which uses only the local-level information and the Laplacian weighted centrality ( $C_L$ ) proposed in [29]. The  $C_L$  of a node in a weighted network graph (interaction strengths are given as weights) defined as the relative drop of Laplacian energy due to deactivation of all interactions of that animal with others in the group. The study [29] derived an approach to compute it by using intermediate-level connectivity information, i.e., the interactions of an animal with first and second NNs. The  $C_L$  formula is explained briefly below and we refer to [29] for the derivation and more details about  $C_L$ .

$$C_L(i) = 4C_W^{c(i)} + 2C_W^{m(i)} + 2C_W^{h(i)}, \quad \text{for } i = 1, 2, \dots, N, \quad (3)$$

where  $S(i)$  contains the set of  $1^{st}$  NNs of the  $i^{th}$  animal and  $W_{i,j}$  is the weight of the interaction between  $i$  and  $j$ ;  $C_W^{c(i)} = \sum_{j \in S(i)} W_{i,j}^2$  is known as the number of closed 2 step interactions containing the animal  $i$  (i.e., from  $i$  to  $j$  and  $j$  to  $i$ );  $C_W^{h(i)} = \sum_{j \in S(i)} \left( \sum_{l \in \{S(j)-i\}} W_{i,j} W_{j,l} \right)$  is the 2 step interactions containing the animal  $i$  as an edge vertex of the path (i.e.,  $i$  to  $j$  and  $j$  to  $l$ , where  $l$ s are the  $1^{st}$  NNs of entries of the set  $S(i)$ );  $C_W^{m(i)} = \sum_{j,l \in S(i), j \neq l} W_{j,i} W_{i,l}$ , represents the 2 step interactions containing the animal  $i$  in the middle of the path (i.e.,  $j$  to  $i$  and then  $i$  to  $l$ ).

#### D. Gaussian Mixture Models

In order to get an overall idea about the variation in animal importance of a group, the joint distribution of  $AI_m$  of all members was derived using the Gaussian mixture model (GMM) approach. This approach was based on the assumption that the  $AI_m$  of each animal has a Gaussian probability distribution (hereafter, termed PDF). In the present study, the PDF of the  $AI_m$  of an animal represents the distribution of  $AI_m$  values collected over a period of time. GMM is a parametric probability density function of a set of PDFs and is computed as their weighted sum. This can also be considered as a hybrid version of a set of PDFs, and provides not only a smooth overall distribution, but also its components unveil the details about the multi-model nature of the density.

Suppose the mean and covariance of a  $D$ -dimensional continuous Gaussian random variable  $X$  is  $\mu_{(1 \times D)}$  and  $\Sigma$ , respectively. Theoretically, the PDF of  $X$  is denoted as  $X \sim \mathcal{N}(\mu, \Sigma)$  and written as:

$$f(X|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)}.$$

Then, the GMM of a set of  $m$  such random variables is written as:

$$F(X|\Phi) = \sum_{i=1}^m w_i f(X|\mu_i, \Sigma_i),$$

where the GMM model weights,  $w_i$ 's satisfy  $\sum_{i=1}^m w_i = 1$  and  $\Phi = \{w_i, \mu_i, \Sigma_i\}$ ,  $i = 1, \dots, m$ , is the set of parameters in GMM model that has to be estimated.

To estimate  $\Phi$ , the iterative expectation-maximization (EM) and maximum A-posterior estimation (MAE) techniques are commonly used. Depending on the different characteristics of the set of PDFs, GMM can have various forms. The covariance,  $\Sigma$ , could be similar for all components in some cases, while it is possible to use a full rank or diagonal covariance matrix. In our evaluation, a full covariance matrix method is used and the EM technique is used to estimate the parameters. Moreover, the GMM model configurations depend on the available data as well as the application. We refer to [30] for more details about GMM.

For instance, if  $X_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ ,  $X_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$ , and  $X_3 \sim \mathcal{N}(\mu_3, \Sigma_3)$  with  $D = 1$ , then the GMM of  $X_1, X_2$ , and  $X_3$  (say  $X_4$ ) can be written as:

$$F(X_4|\Phi_4) = w_1 \mathcal{N}(\mu_1, \Sigma_1) + w_2 \mathcal{N}(\mu_2, \Sigma_2) + w_3 \mathcal{N}(\mu_3, \Sigma_3),$$

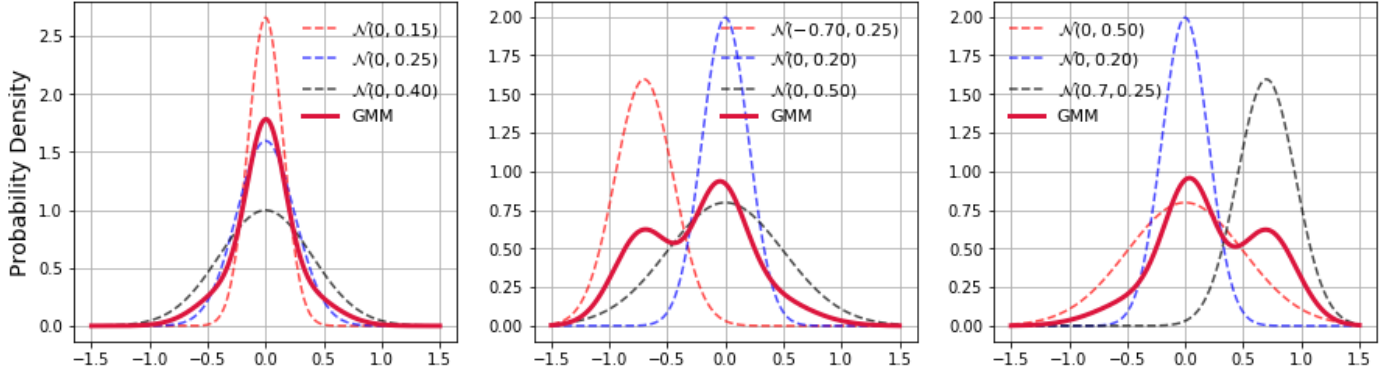


Fig. 3: Variation of the shape of GMM in response to the significant change of the mean of PDFs. First one for similar  $\mu$ s and different  $\sigma$ s. The second and third graphs represent the change of the shape of GMM when one PDF is shifted to the left and right of the other PDFs two, which have similar  $\mu$  and different  $\sigma$ s, respectively.

where  $\Phi_4 = \{(w_1, w_2, w_3), \mu_4, \Sigma_4\}$  and  $w_1 + w_2 + w_3 = 1$ .

Figure 3 represents the variation of the shape of the GMM of three different Gaussian random variables  $X_1, X_2$ , and  $X_3$  (where we assume  $D = 1$  and  $\Sigma_i = \sigma_i$  for  $i = 1, 2, 3$ ) in response to the change in their PDFs for three different cases; 1) similar means with different variances ( $\mu_1 = \mu_2 = \mu_3$ ) (left), 2) variance of  $X_1, X_2$ , and  $X_3$  are different, and  $X_2$  and  $X_3$  have similar means, but  $X_1$ 's mean diverged to the left of the mean of  $X_2$  and  $X_3$  ( $\mu_1 < \mu_2 = \mu_3$ ) (middle), and 3) the opposite case of the second case, but the change happens with the variable  $X_3$  instead of  $X_1$  ( $\mu_1 = \mu_2 < \mu_3$ ) (right). In each case, the variance of  $X_1, X_2$  and  $X_3$  are different from each other, i.e.,  $\sigma_1 \neq \sigma_2 \neq \sigma_3$ . So, the GMM represents a multi-model nature (in this case it is a bi-model because there are two-packs only) mainly when the means of PDFs are significantly different.

When considering the  $AI_m$  of an arbitrary animal as a random variable similar to  $X$  explained above with  $D = 1$ , the PDF of the  $AI_m$  of that animal can be written as  $f(X|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$ , i.e.,  $X \sim \mathcal{N}(\mu, \sigma)$ . Then the GMM of  $AI_m$  represents the joint distribution of the  $AI_m$  of the group.

#### E. Geo-Location Mobility Data

Global positioning system location data of 33 ( $N$ ) cows (a herd) were collected over a 24 h period for 5 consecutive days ( $T$ ) at a Teagasc research dairy farm, in Moorepark, Ireland. All cows were part of a study on robotic milking in grazing production systems [31]. The data were partitioned into non-overlapping time windows each of 10 min ( $t$ ) duration (i.e.,  $T_w = 720$  windows for the entire duration). For each animal, maximum three locations were observed in every 10 min, so that the average location of each cow was computed within each window. Linear interpolation was used to compute missing observations. The cows grazed different paddocks during the 5-day period with varying distances from the milking parlor (Figure 4). All cows visited the milking station at most twice daily and moved to a new paddock after each milking.

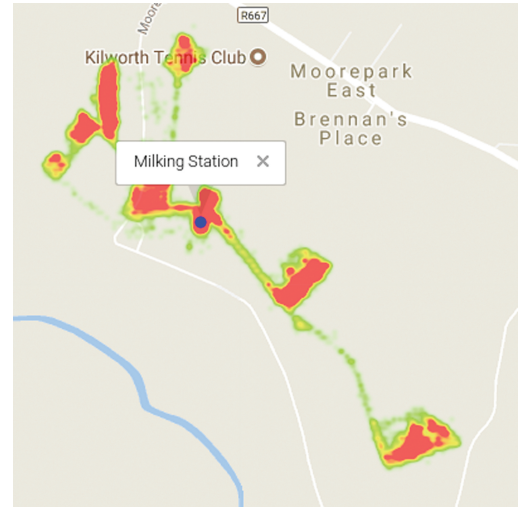


Fig. 4: Heat map of the global positioning system locations of the cows at the research farm in 5 days period; the blue dot is the milking station.

#### IV. RESULTS

The applicability of the mathematical procedures is discussed by using a GPS mobility dataset of a dairy herd explained in Section III. First, the cattle social interaction range is derived using the dataset and it is then validated based on the variability of network entropy. The variability of social affiliation of dairy cattle is discussed next. Finally, the entropy-based  $AI_m$  is computed and its validity is compared with the importance evaluated using the direct degree and Laplacian centrality measures. MySQL database was used for efficient data storage and access and Python software was used for data analysis and the simulation study. More specifically, the Python packages *networkx* and *scikit-learn* were used for building the network graph and computing GMM models, respectively.

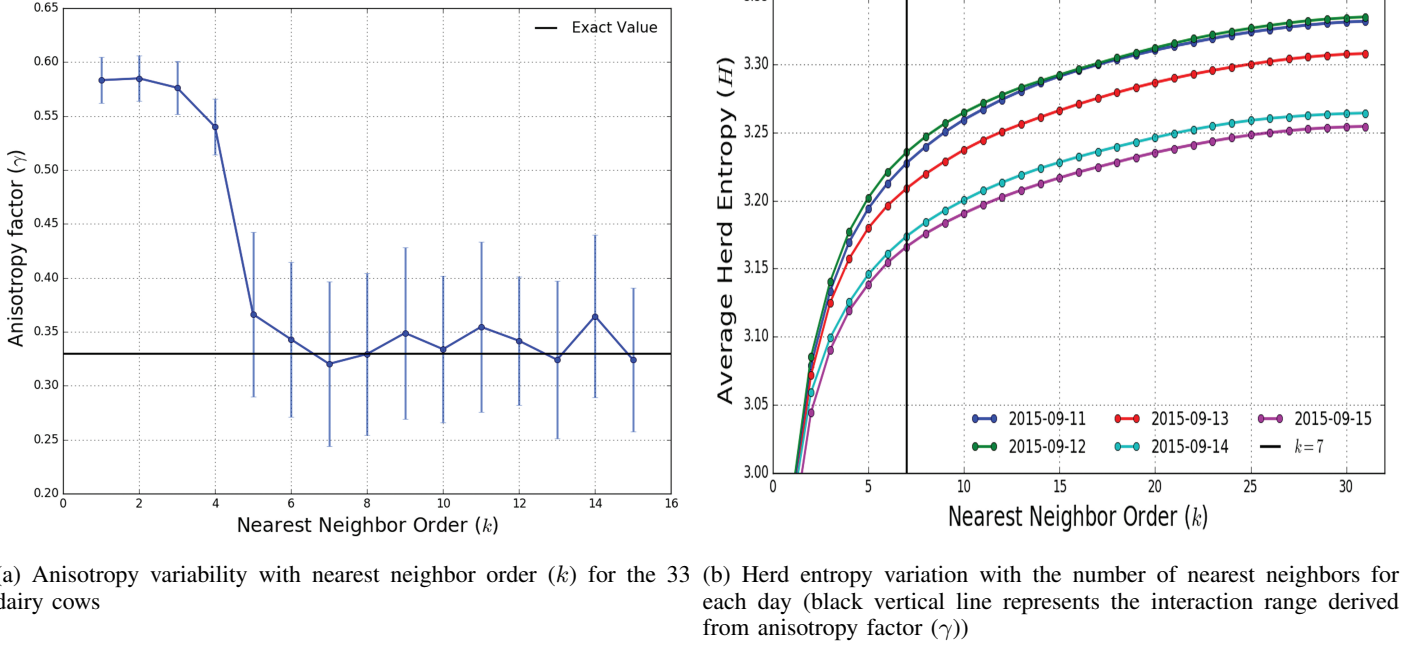


Fig. 5: Derivation of dairy cattle social interaction range ( $k_r$ )

#### A. Interaction Range of the Pasture-based Dairy Animals

The anisotropy factor ( $\gamma$ ) revealed a clear relationship with the herd center velocity ( $V$ ) along the direction of the eigenvector. This corresponded to the smallest eigenvalue of the average projection matrix ( $M'$ ) (Figure 5a). The  $\gamma$  decayed with increasing NN order ( $k$ ) and reached 1/3 (exact value or isotropic point) near a  $k$  value of 7. Thus, the NNs above the 7<sup>th</sup> order NN were isotropically distributed around a focal animal and did not strongly interact with it. Therefore, the value of interaction range ( $k_r$ ) was selected as 7 ( $k_r = 7$ ). In other words, the strength of the interactions of a focal cow with the first 7 NNs was considerably higher compared to the interactions beyond the 7<sup>th</sup> NN.

The variation in average herd entropy with the spatial distribution of NNs over the 5 days of the study is shown in Figure 5b. Although the herd entropy increased with the NN order (i.e.,  $k$ ), it did so at a declining rate. This is because entropy was calculated based on the weighted degree centrality,  $C_W$ . The weights of interactions reduced with the increasing order of NN. Therefore, the contribution to the increment of herd entropy,  $H(G)$  from the interactions with the animals beyond the  $k_r^{th}$  NN was minimal (on average  $\leq 0.01$ ). This indicates that the interaction range of our pasture-based dairy cows is 7.

To be more precise about the interaction range derived from Figure 5a, the effect of the interaction range on the herd entropy was examined. The herd entropy characterizes the information about the connectivity of the herd. Hence the broader interaction range will increase the herd connectivity

and so will the herd entropy. However, the interaction strengths weakened as the interaction range increased and consequently, contributed to a decline in the increasing rate of herd entropy. Therefore, the optimal interaction range (i.e., number of NNs) which does not make any significant contribution to improve the herd entropy should be the optimal interaction range (Figure 5b). Although having a broader interaction range brings some disadvantages such as higher competition for food and space and risk of falling sick [32], there are some benefits as well, especially in graph-based SNA and also for corporate defending against predator attacks.

#### B. Time-Evolving Interactions with NN Frequency Matrix

The frequency of each cow being within the interaction range of every other cow over the 5-day period of the study is represented in Figure 6, which can be read similar to reading a normal square matrix. For instance, considering the cow index (ID) 10, the corresponding row represents the frequencies of interactivity of all other cow IDs with the cow ID 10, while the column depicts the frequencies of interactivity of cow ID 10 with the remaining cow IDs, including itself. Since the NN frequency matrix is not a symmetrical matrix, the entries of a particular row were not always exactly similar to the entries of the corresponding column. As an example, the interactivity of the cow ID 10 to the cow ID 9 was greater than 70%, whereas it was less than 60% for the cow ID 9 being interactive with the cow ID 10 over the 5-day period of the study.

The variability of frequency fluctuations provided a clear illustration of the intensity of inter-animal interactions (social



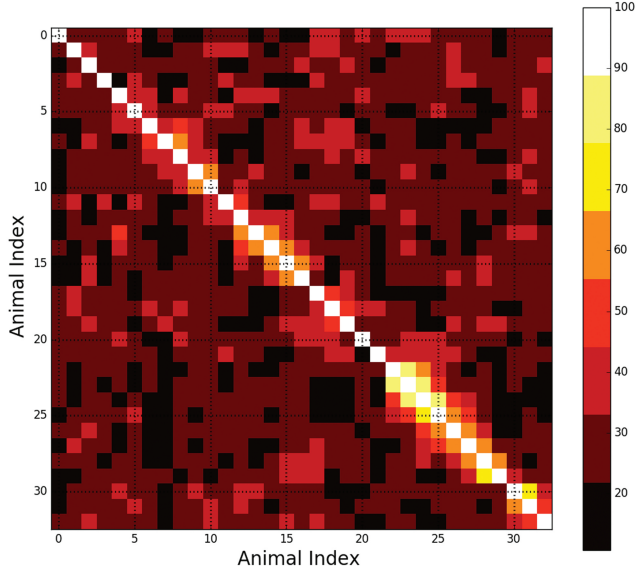


Fig. 6: The frequency of being individual's interaction range as a percentage of the total number of time-windows (i.e., 720) over the 5 days period (the size of a time-window was 10 min and the frequency percentages were computed as  $(f/720) \times 100$ ).

affiliation) in order to identify specifically the most and least interactive cows. In general, most of the frequency values were below 40% (i.e., less than 30 out of 120 h), but few cow ID pairs represented greater than 50% values in Figure 6. For instance, the interactivities of the cow ID pairs 7-8, 13-15, and 29-31 were between 60% and 70% and it was greater than 80% for the cow ID pairs 22-23, 23-24, and 24-25.

Since the highly interactive cows are easily recognizable in Figure 6, this matrix could be used to select especially the most (or least) cohesive cows in a herd. Hence, this matrix could be useful in characterizing herd-level behaviors such as the cows which are at high (or low) risk of getting viral infections, the leading disease carriers, sexually active cows, and the cows which are highly likely to synchronize their activities. Such information would be useful to manage herds more efficiently in different farm operations such as maintaining optimum animal well-being and training cows for voluntary participation in milking. In SNA point of view, sampling animals form larger group to form network graphs.

### C. Animal Importance to the Herd

The importance ( $AI_m$ ) of the 33 study cows from the weighted degree centrality and Laplacian centrality measures ( $C_W$  and  $C_L$ ) and herd entropy variation over the 5-day study period are shown in Figure 7. The  $AI_m$  based on the  $C_W$  and  $C_L$  measures were more or less similar and fluctuated more with greater variability compared to the herd entropy-based  $AI_m$ . The  $C_L$ -based  $AI_m$ s were more stable than the  $AI_m$  computed using  $C_W$  because the confidence limits of the  $C_L$ -based  $AI_m$  were slightly tighter than the confidence limits

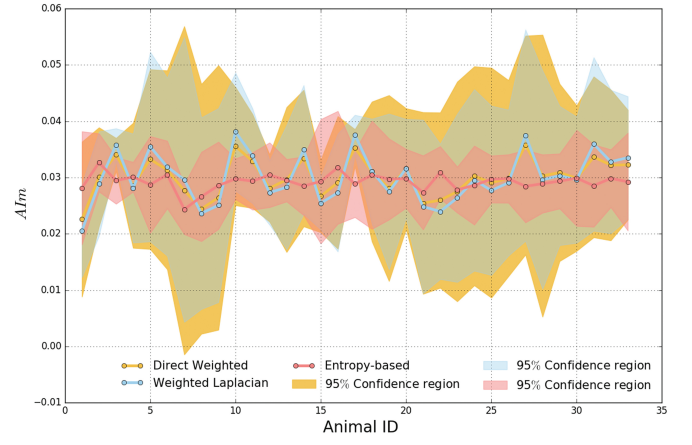


Fig. 7: The  $AI_m$ s of 33 study cows and their variability from the direct weighted centrality ( $C_W$ ), weighted Laplacian centrality ( $C_L$ ), and herd entropy-based approach.

of the  $AI_m$  computed using  $C_W$ . On average, the confidence limits of the herd entropy-based  $AI_m$  was considerably tighter compared to  $C_W$  and  $C_L$  measures based  $AI_m$ s. Therefore, the stability of the  $AI_m$  computed using the entropy-based measure is greater than the other two measures.

The degree centrality is itself a measure of animal importance to the herd and can be computed mainly using the direct and weighted degree centrality ( $C_W$ ) measures.  $C_W$  takes into account the heterogeneous nature of social connectivity based on their interaction strengths, but the direct degree centrality assumes that all interactions have the same strength. Hence, the  $C_W$  is a commonly used reliable measure for quantifying animal importance in SNA. Moreover, the accuracy of  $C_W$ -based  $AI_m$  could even be improved by deriving the interaction strengths (weights) from well-defined methods [21] rather than inter-animal distance. Although  $C_W$  performs well in quantifying animal importance, based on local-level connectivity information, it has been reported that the reliability of animal importance measure could be improved further by increasing the range of connectivity information based on the  $C_L$  measure defined by [29].

However,  $C_L$  uses only the intermediate-level information (i.e., the connectivity data of each animal associated with the single (direct) and two-step interactions). Therefore, to improve the robustness of  $AI_m$ , the present study considered the global-level information (i.e., herd entropy) to compute animal importance because highly reliable  $AI_m$  is necessary for detecting animals which have significant influence, such as dominant cows and group leaders. That is why the uncertainty of the entropy-based  $AI_m$  of the 33 study cows was considerably smaller (with a narrower confidence interval) compared to  $C_W$  and  $C_L$  in Figure 7. Thus, this guarantees that the entropy-based  $AI_m$  characterizes the individual cow influence to the herd more precisely than the existing measures.

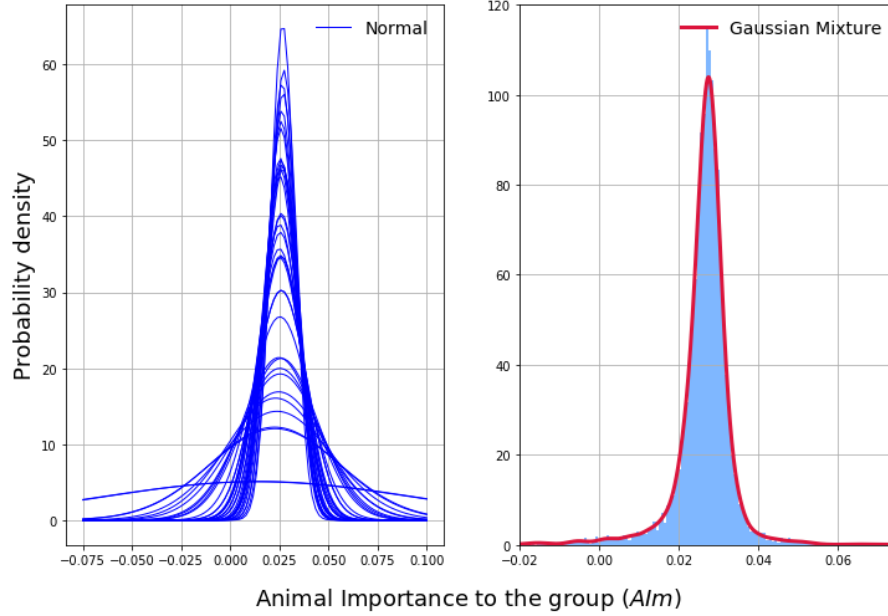


Fig. 8: The individual PDF of  $AIm$  of normal herd animals(left) and their joint PDF (right).

## V. DISCUSSION

The ‘datafication’ of modern-day dairy and other agricultural production systems, through the widespread adoption of sensors [33], facilitates the development of novel metrics and algorithms to detect individual animals that may require particular attention. This is especially true with the proliferation of these technologies which, because of their ever-reducing costs are now being used on an individual animal basis. It was our hypothesis that a greater exploitation of the data from all individuals in a group could provide more information than the analysis of the data relating to just a single animal in the process of evaluating  $AIm$ . In this section, we discuss two use cases of  $AIm$  and directions for further studies based on the limitations and benefits of the  $AIm$  approach.

### A. Animal Mobility Simulation to Identify Estrus and Sick (or Injured) Animals

In order to demonstrate how to use  $AIm$  measure in a real-world application, an example is illustrated regarding the detection of sick and estrus cows in a herd. The illustration is, however, based on the simulated mobility data (using the *pymobility* Python package [34]) because the dataset used above did not have any estrus or sick cows and was too small, and was collected only over a short period of time.

In the simulation study, the behavior of sick and estrus cows was compared to normal contemporaries in a herd based on their variability in the  $AIm$  metric. Two steps were taken in carrying out the simulation study. First, the mobility patterns of the sick and estrus cows relative to the normal cows (i.e., non-estrus and healthy) were simulated separately over a time period by using a random-way-point mobility pattern [34]. This approach is commonly used to simulate animal movements within a given region, depending on their velocity and waiting

time (time spent at a position before making the next random movement). Therefore, variability in the characteristics of walking area, speed and waiting time were considered as the signs (parameters) in simulating mobility of sick (and estrus) cows. Second, to represent the variability in  $AIm$ , the PDF of  $AIm$  was computed for each sick animal, each animal in estrus, and each normal animal. Also, to illustrate the group-level variation of  $AIm$  of sick and estrus herds with respect to normal herds, the joint PDF of the individual PDF were derived.

Three herds (i.e., normal, sick, and estrus animals) each with 40 cows were simulated separately within a  $60 \times 30m$  (rectangular paddocks are preferred with 2:1 width and depth ratio<sup>1</sup>) paddock over 1500 time windows ( $\sim 10$  days). In sick and estrus herds, 35 of the cows were assumed to be normal and the rest were assumed to be either sick or in estrus. In this simulation, we did not consider sick and estrus animals together in the same herd. This is because there might be a cow in both conditions and it might be difficult to differentiate between such conditions in the same animal.

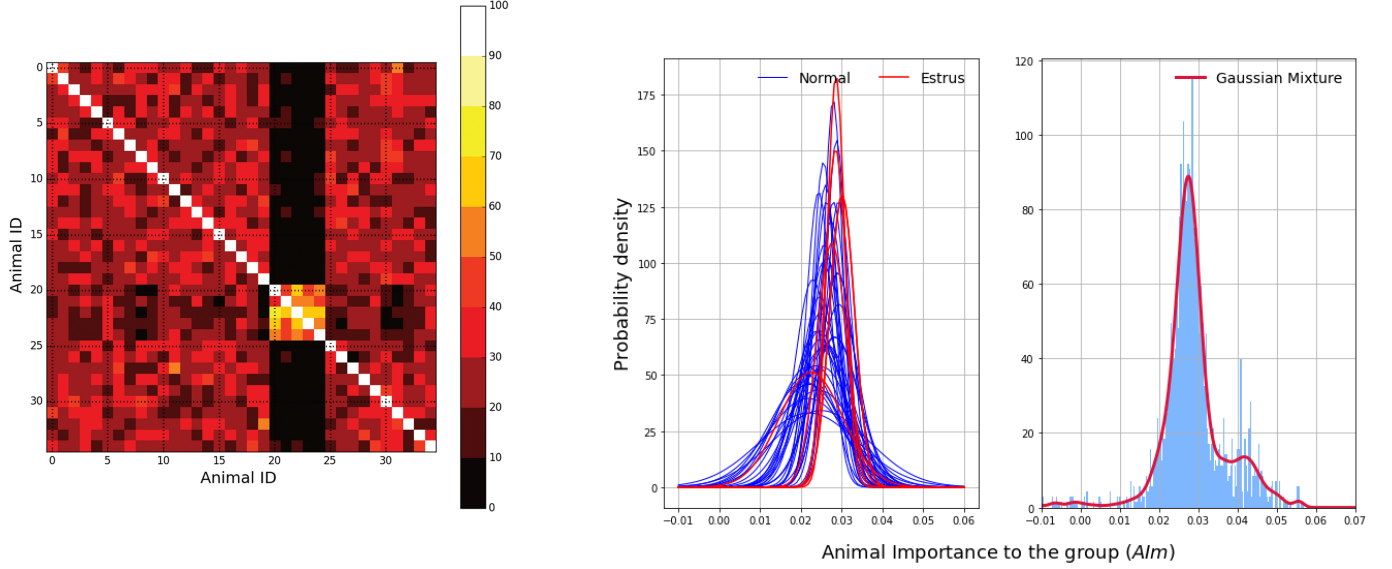
According to the literature, the average cow speed is 2-3 km/hour and it can be increased up to 4.5 km/hour in properly maintained farms. However, these ranges are affected by many factors such as paddock and track design, weather conditions, and distractions<sup>1 2</sup>. In optimal conditions, cows lie down generally for 14 hours per day ( $\approx 30min/hour$ ) and on average, a normal cow gets up 16 times a day. The resting time varies for many reasons such as age, heat cycle, health, weather, herd size, and the housing and paddock conditions<sup>3</sup>. Therefore, in

<sup>1</sup><https://www.teagasc.ie/media/website/animals/dairy/GrazingInfrastructure.pdf>

<sup>2</sup><https://www.dairynz.co.nz/media/214237/Understanding-cow-movement.pdf>

<sup>3</sup><http://www.milkproduction.com/Library/Scientific-articles/Housing/Cow-comfort-9/>





(a) The intensity of interactivity of cows in estrus (IDs 20-24) with the non-estrus animals. (b) The individual PDF of  $AIm$  of the estrus animals with their normal herd animal (left) and their joint PDF (right).

Fig. 9: The variation of social affiliation and importance of the estrus and normal cows.

the simulation study, cow velocity and the waiting time at a position were assumed to vary uniformly in the ranges of 0.0 - 3 km/hour and 0 - 30 min/hour, respectively. Also, we assumed that normal cows move over the entire paddock at their own peace.

As sick cows move slowly and take more time to make the next move, they have long resting (waiting and lying) time. Therefore, to simulate the mobility pattern of a sick cow, the region covered and the velocity range were reduced by half and the waiting time was doubled compared to a normal animal. This procedure was repeated for all time windows and the PDFs of  $AIm$  for each animal were quantified. The same process was followed for the herd with cows in estrus by changing the mobility pattern of cows in estrus; since the level of activity intensifies when cows are in estrus [35], the velocity range was doubled and the waiting time was halved. These cows were allowed to move in the entire paddock the same as the normal cows in each iteration.

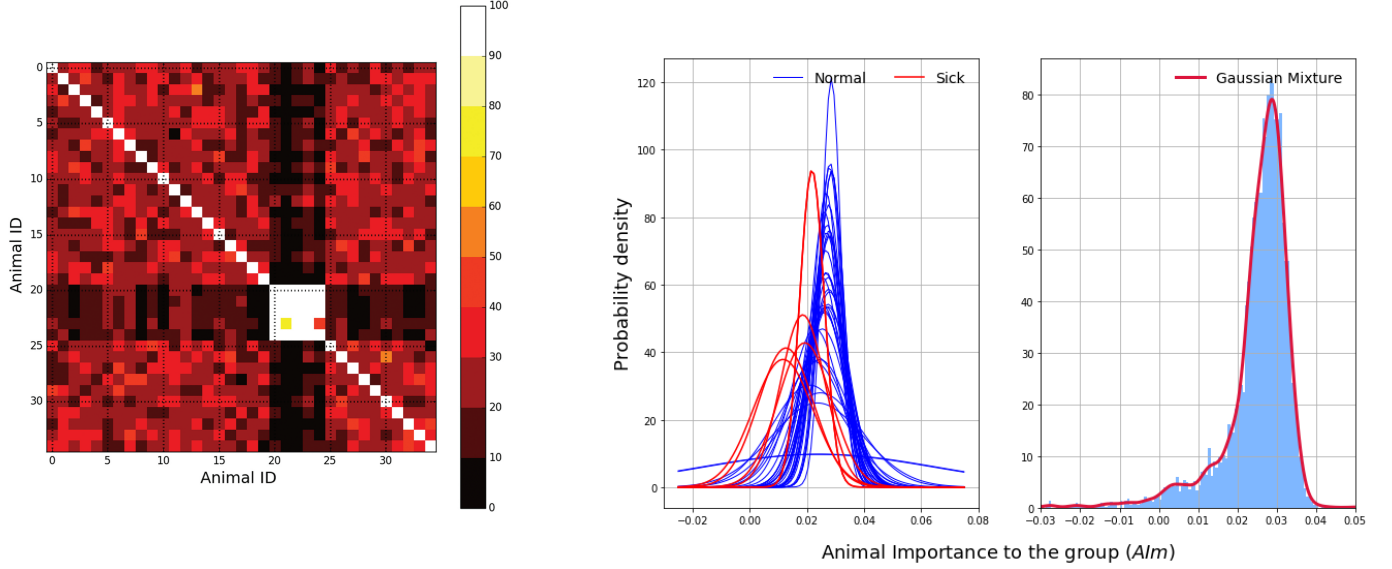
1) *Normal Herd* : Figure 8 represents the PDF of  $AIm$  of individual animals (left) and their joint PDF (right). In general, all PDFs are distributed around a common mean (approximately 0.025) with different shapes (variances). The corresponding mixture (joint) PDF represents the overall distribution of  $AIm$  of the herd and does not indicate any multi-model nature as there is only one peak similar to the theoretical example explained (first case) in Figure 3. Therefore, any deviation from this behavior would be an early indication of herd approaching to an unusual behavior and thus needing attention. The next two sections illustrate the nature of such

deviation when there are sick and estrus animals in the herd.

2) *Identifying Cows Approaching or in Estrus*: Figure 9a and 9b present the simulated outcomes of the variability in interactivity and  $AIm$  (as PDFs) of a mixture of estrus and non-estrus cows, respectively. In general, the intensity of interactivity between the non-estrus cows ranged between 30-50 % (IDs 20-24 represent the estrus cows; Figure 9a) and, on average, the intensity of interactivity among only the cows in estrus was greater than 70%, but was less than 10% between the estrus and non-estrus cows (columns of the estrus cows; Figure 9a). On the other hand, the intensity of interactivity of normal cows to the estrus cows (rows of the estrus cows) was less than 30% (Figure 9a). The mean values of the PDFs of non-estrus cows were approximately similar (i.e., the PDFs distributed around a similar mean value), but their variance were different (Figure 9b).

The PDFs of the estrus cows were shifted towards the right of the PDFs of the normal cows and denser (i.e., less variance, high kurtosis) than normal cows. Consequently, the joint of these PDFs represents a bi-model behavior as there are two peaks in the distribution. In comparison to the GMM of normal herd given in Figure 8, the GMM is expanded towards the right with a peak value. This is due to the cows in estrus having higher  $AIm$  and validates the theoretical illustration (case 3) given in Figure 3.

The mobility patterns of the sexually active cows are generally different from the normal cows due to their greater tendency to join groups of cows that are also in estrus, and thus have less resting time, as well as ending up walking together.



(a) The intensity of interactivity of sick cows (IDs 20-24) cows (b) The individual PDF of  $AIm$  of the sick animals with their normal herd animals (left) and their joint PDF (right).

Fig. 10: The variation of social affiliation and importance of the sick and healthy cows.

Consequently, the measured interactivity among cows in estrus is greater than the interactivity between cows in estrus and non-estrus. Hence, the simulated intensity of interactivity between the cows in estrus was greater compared to the intensity of interactivity between estrus and normal cows in Figure 9a. Consequently, cows in estrus become the most important (i.e., highly cohesive) members of the herd more frequently than normal cows. As a result, less variance and higher mean of their  $AIm$  is detected. Therefore, the PDFs of  $AIm$  of cows in estrus become more dense compared to that of the normal cows (Figure 9b) and also, the joint PDF of  $AIm$ s is more expanded towards the right compared to normal the herd (Figure 8).

The failure of detecting cows in estrus in herds not adopting fixed-time artificial insemination has negative repercussions for farm profit owing to extended calving intervals as well as possibly the cost of semen inseminated at the incorrect time of the estrus cycle [33]. [36] noted that herd behavior should be closely monitored as some of the well-documented signs of estrus may not exclusively be exhibited by the cows in estrus. Among the different herd behaviors which have been categorized as primary and secondary signs of estrus [37], observing the change of cow mobility patterns can be useful as a behavioral sign for generating early alerts about the cows approaching or in estrus. Therefore, the matrix  $A$  and  $AIm$  can be used in detecting early estrus signs.

3) *Identifying Sick or Injured Cows:* Figure 10a depicts the variability of intensity of interactivity for a mixture of sick and normal cows. The variability in the intensity of interactivity of sick cows (20-24 rows and columns) is clearly highlighted in

Figure 10a evident. The intensity of interactivity of sick cows with the herd was below 40%, while among the sick cows it was greater than 80%. The PDFs of the sick animals shifted towards the left of the mean value of the PDFs of normal cows (Figure 10b), and also their mean values were different to the mean values of the PDFs of normal cows. As a result, the overall distribution of the  $AIm$  of the herd is expanded to the left side with a peak, i.e., bi-model behavior with left tailed distribution.

Due to the fact that sick cows are reluctant, or have difficulty in moving, they might not always be able to follow the herd. They may become isolated from the herd or lag behind the herd. Consequently, their mobility patterns may be different (or diverge) from their peers in the herd. This divergence can therefore be easily used to differentiate those animals. In SNA, this divergence of mobility can be accounted for as a decline in their interactivity with the herd. Hence, highly interactive cows will generally not be the sick cows. Also, they will have fewer chances of being as highly interactive members to other herd mates. So, the intensity of interactivity of sick cows with the herd generally becomes low (below 30% in Figure 10a). Moreover, sick cows may gather as a sub-group, so that they are counted as being highly interactive with each other while computing the NN frequencies. That is why Figure 10a represents the greater intensity of interactivity (greater than 80%) values for sick cows. But in practice, this does not imply that they are exclusively associated to each other. As sick cows experience different mobility patterns and are less interactive within the herd over time, their  $AIm$  becomes

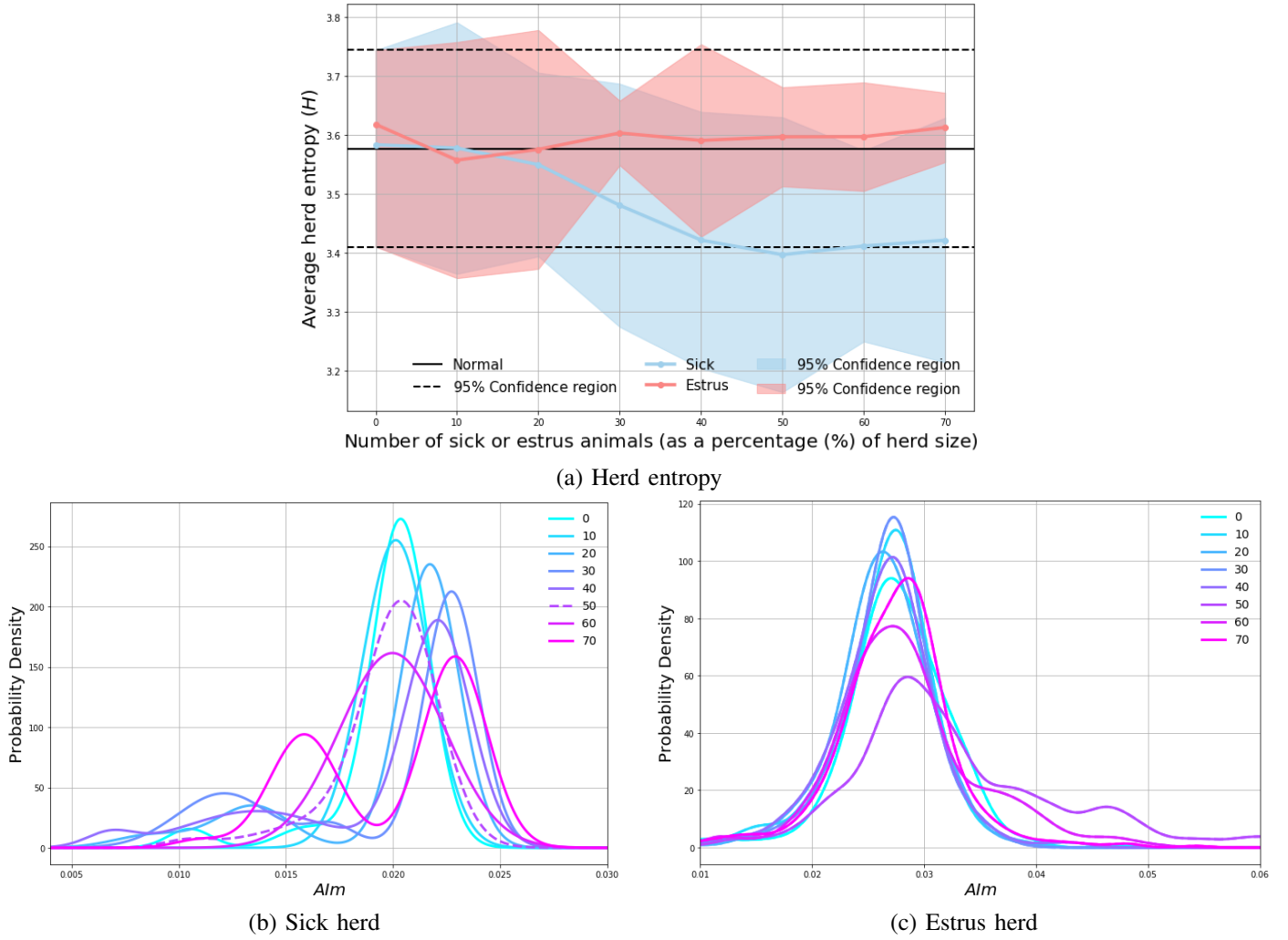


Fig. 11: The herd entropy variation with the sick and estrus animals accounted as a percentage of herd size (a) and corresponding variation of GMM for the sick (b) and estrus (c) herds.

less and varies more compared to normal cows because  $AIm$  depends on the number of interactions and their weight. Hence, the mean values of their PDFs of  $AIm$  were less and different to the healthy cows in Figure 10b. Consequently, the joint PDF of  $AIm$  is expanded more towards the left, compared to the normal herd joint PDF given in Figure 3. This validates the theoretical concept explained under the case 2 in Figure 3.

Early detection of health-related ailments is extremely important to prevent erosion of herd profit, but also to minimize the impact on animal welfare. In the early attempts of animal health assessment studies, animal mobility was commonly used to identify animal health issues [38]. Moreover, [39] stated that automatically recording mobility behaviors facilitates the derivation of valuable animal health alerts. Apart from the cow mobility pattern analysis, lying and standing behaviors [40], body temperature and visual signs [40], and variation in hunger [41] have also been used to evaluate animal health

status.

4) *Herd Behavior with Increasing Sick and Estrus Animals:* To explore the herd behavior as more animals become sick in the herd, under the same settings, individual  $AIm$  and herd entropy were computed by varying the number of sick animals as a percentage of the herd (0% to 70% prevalence within the herd). To represent variability in  $AIm$  and herd entropy, the 95% confidence limits were also computed. It was assumed to remain that animals falling sick were sick for the rest of the simulation process. The same procedure was repeated with increasing number of estrus animals in the herd. As cows are in estrus for only a few hours, we did not however assume that estrus animals stay remain in estrus for the rest of the simulation.

The variability in average herd entropy with increasing number of sick and estrus animals (as a percentage of herd size) in the simulated herd is shown in Figure 11(a). The

herd entropy of the normal herd (i.e., 0% sick or estrus cows) was  $3.58 \pm 0.13$ . As the sickness spread in the herd, the average herd entropy decreased and the associated confidence interval widened. With increasing number of estrus cows, the herd entropy slightly increased compared to the normal herd entropy, the confidence limits of the herd entropy varied within the confidence limits observed for the normal herd and also became narrower. Furthermore, as a consequence of increasing the number of cows in estrus in the herd, there will be many high importance animals. Therefore, the variability in herd entropy will be smaller (i.e., more stable) and so will the individual  $AI_m$ . This is the reason for obtaining narrower confidence interval for the herd entropy as more cows came to estrus in Figure 11(a). Hence, examining the variability in herd entropy and individual  $AI_m$  over a period of time can also be used to generate early alerts about the cows in or approaching estrus.

Figures 11(b) and 11(c) depict the variability in GMM of  $AI_m$  as the number of sick and estrus animals increased in the herd, respectively. When the prevalence of sick animals in the herd increased, the GMM spread widely towards the left of the mean of the normal herd (0% sick animals) and also, the bi-model nature became more apparent. Similarly, when more animal come to estrus, the GMM represented a tendency of spreading towards the right from the mean of the non-estrus case (0% case). The possibility of dividing the herd into sub-groups is increasing with increasing sick/estrus animals in the herd. That is why the multi-model nature of GMM is increasing in both Figures 11(b) and 11(c). Moreover, the disperse (uncertainty) of GMMs is higher with the prevalence of sickness in the herd than the increment of cows in estrus.

Therefore, exploring the most prominent features in the NN frequency matrix and the variability (or deviations) in herd entropy, individual PDF of  $AI_m$ , as well as the GMM of  $AI_m$  over a time scale, can be used to detect the presence of sick cows which may require closer attention.

### B. Directions for Future Studies

The  $AI_m$  cannot, however, capture time-variant properties of dynamic networks because  $C_W$  used to derive  $AI_m$  and most of the other graph-theoretic measures have been defined for static graphs [40], [18]. So, novel measures for dynamic network graphs are highly desirable in social network analysis. Moreover, animal interactions are heterogeneous in nature, and also there are various factors which influence the generating strong interactions such as gender, age, and parental attraction. So, computing interaction weights by taking into account such factors would definitely contribute to capture more precise social behavioral characteristics. Moreover, taking into consideration more mobility parameters such as the probability for deciding the next movement direction and using the current location information when deciding the next location, would help to simulate more realistic mobility data and so would the social behaviors.

In modern-day data analytic applications, interoperability of heterogeneous datasets collected across a geographically distributed data source is one of the critical issues. The reason

for that is most of the technologies currently in use, notably in agri-sector, operate in isolation as they are incapable of communicating with each other. Consequently, the importance of collected data from those technologies is significantly under-utilized. To make full use of such data, it is necessary to explore the inter-relationships between such datasets to conduct a more comprehensive analysis. The use of  $AI_m$  facilitates incorporating location-based mobility behaviors with a variety of other information sources. For instance, exploring the impact of cow mobility on milk and pasture quality, feed intake, and nutrient deficiencies. Therefore, exploring the interrelationship of  $AI_m$  with other farming variables would be another exciting extension of the present study that would help alleviating the barriers of performing cooperative analytics with heterogeneous datasets.

Additionally,  $AI_m$  evaluating process could also be used in other applications such as wireless sensor network (WSN) and distributed data analytic platforms, which are currently used in real-time decision making. In such applications, optimizing the computational, communication, and energy requirements are critical in order to improve the responsiveness and timeliness of the system. In this case, identifying least and highest  $AI_m$  nodes can make significant contribution. For instance, while the high  $AI_m$  nodes are vital in selecting nodes (gateways) to improve the effectiveness of data communication in the network, least  $AI_m$  nodes would be the best nodes for offloading computations and sharing resource requirements. Moreover, in federated learning, which is a distributed machine learning framework, a machine learning model is trained across a large number of data sources, and the model updates are aggregated at a coordinating unit to compute the final model updates. Those updates are then sent back to each data source to make inferences. So, the  $AI_m$  metric can be used to select the coordinating unit as the data source which has the highest  $AI_m$  as such a node can effectively communicate with other data sources. Therefore, the responsiveness of the learning system can effectively be increased, optimizing the resource consumption.

## VI. CONCLUSIONS

The NN-based graph analytic techniques were used in the present study to evaluate animal importance to the group, combining animal- and group-level information. Meanwhile, animal interaction range and NN frequency matrix were derived to support the  $AI_m$  evaluation process. Based on the observations from the cow sample population in the present study, the most common interaction range of grazing dairy cows is 7 and the NN-frequency matrix gives an overview regarding the social affiliation of animals in the herd. The  $AI_m$  metric derived based on the herd entropy variation to quantify  $AI_m$  performed well compared to the already existing  $AI_m$  measures, which are based on the degree and Laplacian centrality measures. Using simulated data, the intensity of cohesion in cows in estrus was greater among themselves than their cohesion to the herd. The sick cows demonstrated considerably less intensity of cohesion to the herd than healthy cows and that was even smaller compared to the cows in estrus.

The PDFs of cows in estrus were shifted to the right of the mean of the PDFs of non-estrus cows, which exhibited nearly a similar mean, but differences existed in their variances. Also, cows in estrus had smaller variances (i.e., more dense PDFs) than their non-estrus contemporaries. The mean values of the PDFs of sick cows were not comparable to each other and were less than the healthy cows. Moreover, while the joint PDF of  $AI_m$  in sick herd represented a bi-model nature and expanded towards the left compared to the joint PDF of  $AI_m$  of a healthy herd. The joint PDF of on estrus herd showed completely an opposite behavior to the sick herd. Results from this study therefore suggests that the NN frequency matrix and entropy-based animal importance metric can be used to generate early alerts about the deviations of their social behaviors and then to derive useful information.

### ACKNOWLEDGEMENTS

This research was supported by the Science Foundation Ireland (SFI) project "PrecisionDairy (ID: 13/1A/1977)" as well as a research grant from Science Foundation Ireland and the Department of Agriculture, Food and Marine on behalf of the Government of Ireland under the Grant 16/RC/3835 (VistaMilk) and also the Horizon 2020 GenTORE project. The contribution of Diarmuid McSweeney in the collection of the GPS data is also acknowledged. We are also grateful Dr. Bernard Butler for the valuable comments on improving the technical quality of the paper.

### REFERENCES

- [1] N. K. Boyland, "The influence of social networks on welfare and productivity in dairy cattle," <https://ore.exeter.ac.uk/repository/bitstream/handle/10871/19360/BoylandN.pdf?sequence=1&isAllowed=y>, 2018.
- [2] N. K. Boyland, D. T. Mlynski, R. James, L. J. N. Brent, and D. P. Croft, "The social network structure of a dynamic group of dairy cows: from individual to group level patterns," *J. of Applied Animal Behavioral Sciences*, vol. 174, pp. 1–10, 2016.
- [3] E. A. Codling and N. W. Bode, "Copycat dynamics in leaderless animal group navigation," *J. of Moving Ecology*, vol. 2, no. 11, 2014.
- [4] R. N. Handcock, D. L. Swain, G. J. Bishop-Hurley, K. P. Patison, T. W. P. Valencia, P. Corke, and C. J. O'Neill, "Monitoring animal behaviour and environmental interactions using wireless sensor networks, gps collars and satellite remote sensing," *J. of Sensors*, vol. 9, pp. 3586–3600, 2009.
- [5] B. C. Fargas and M. N. Petersen, "Gps-free geolocation using lora in low-power wans," in *IEEE Int. Conf. on Global Internet of Things Summit (GloTS)*, 2017.
- [6] P. Varun, W. Elmannai, and K. Elleithy, "Mobile and wi-fi geo-location using google latitude," in *IEEE Int. Conf. on Computer and Information Technology (WCCIT)*, Sousse, Tunisia, 2013.
- [7] A. Goswami and A. Kumar, "Challenges in the analysis of online social networks: A data collection tool perspective," *J. of Wireless Personal Communications*, vol. 97, no. 3, pp. 4015–4061, 2017.
- [8] E. Pennisi, "Animal cognition: Social animals prove their smarts," *J. of Science*, vol. 312, no. 5781, pp. 1734–1738, 2006.
- [9] T. Slotyasik and Z. Nogalski, "The effect of social hierarchy in a dairy cattle herd on milk yield," *Political J. of Natural Science*, vol. 25, pp. 22–30, 2010.
- [10] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic, "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study," *PANAS*, vol. 105, pp. 1232–1237, 2008.
- [11] A. Cavanga, I. Giardina, A. Orlandi, G. Parisi, and A. Procaccini, "The starflag handbook on collective animal behaviour: Part ii, three-dimensional analysis," *J. of Animal Science*, vol. 76, pp. 238–248, 2008.
- [12] T. Wey, D. T. Blumstein, W. Shen, and F. Jordan, "Social network analysis of animal behaviour: A promising tool for the study of sociality," *J. of Animal Behaviour*, vol. 75, pp. 333–344, 2008.
- [13] B. A. Wood, H. T. Blair, D. I. Gray, P. D. Kemp, P. R. Kenyon, and S. T. Morris, "Agricultural science in the wild: A social network analysis of farmer knowledge exchange," *PLoS ONE*, vol. 9, no. 8, pp. 4015–4061, 2014.
- [14] N. Pinter-Wollman, E. A. Hobson, J. E. Smith, A. J. Edelman, D. Shizuka, S. de Silva, J. S. Waters, S. D. Prager, T. Sasaki, G. Wittemyer, J. Fewell, and D. B. McDonald, "The dynamics of animal social networks: Analytical, conceptual, and theoretical advances," *J. of Behavioural Ecology*, vol. 25, pp. 242–255, 2014.
- [15] X. Qi, R. D. Duval, K. Christensen, E. Fuller, A. Spahiu, Q. Wu, Y. Wu, W. Tang, and C. Zhang, "Terrorist networks, network energy and node removal: A new measure of centrality based on laplacian energy," *J. of Social Networks*, vol. 2, pp. 19–31, 2013.
- [16] X. Ai, "Node importance of ranking of complex networks with entropy variation," *J. of Entropy*, vol. 19, no. 303, 2017.
- [17] K. Park and A. Yilmaz, "A social network analysis approach to analyze road networks," in *ASPRS annual conference*, San Diego, California, 2010.
- [18] K. Lerman, R. Ghosh, and J. H. Kang, "Centrality metric for dynamic networks," in *MLG '10 of the Eighth Workshop on Mining and Learning with Graphs*, New York, NY, USA, 2014, pp. 70–77.
- [19] G. Martino, F. M. Sarti, and F. Panella, "Social network analysis in encouraging role-players in the beef market to take breeding decisions: A methodological study," *Italian J. of Animal Science*, vol. 12, no. 1, 2013.
- [20] S. A. Rands, "Nearest-neighbour clusters as a novel technique for assessing group associations," *The Royal Society Open Science*, vol. 2, 2015.
- [21] L. Candeloro, L. Savini, and A. Conte, "A new weighted degree centrality measure: The application in an animal disease epidemic," *PLoS ONE*, vol. 11, 2016.
- [22] H. Kim and R. Anderson, "Temporal node centrality in complex networks," *J. of Physical Review*, 2012.
- [23] C. C. Dubé, C. Ribble, and D. Kelton, "An analysis of the movement of dairy cattle through 2 large livestock markets in the province of ontario, canada," *J. of Canadian Veterinary*, vol. 51, p. 1254–1260, 2010.
- [24] P. Koene and B. Ipema, "Social networks and welfare in future animal management," *J. of Animals*, vol. 4, no. 1, pp. 93–118, 2014.
- [25] M. Nöremar, N. Håkansson, S. S. Lewerin, A. Lindberg, and A. Jonsson, "Network analysis of cattle and pig movements in sweden: Measures relevant for disease control and risk based surveillance author links open overlay," *J. of Pre. Vet. Med*, vol. 99, pp. 78–90, 2011.
- [26] C. Veness, "Calculate distance, bearing and more between latitude/longitude points," <http://www.movable-type.co.uk/scripts/latlong.html>, 2008.
- [27] G. Neisen, B. Wechsler, and L. Gyax, "Choice of scan-sampling intervals- an example with quantifying neighbours in dairy cows," *J. of Applied Animal Behaviors*, vol. 116, pp. 134–140, 2008.
- [28] K. K. Nambiar, P. K. Varma, and V. Saroch, "An axiomatic definition of shannon's entropy," *J. of Applied Mathematical Letters*, vol. 5, no. 4, pp. 54–46, 1992.
- [29] X. Qi, E. Fuller, Q. Wu, Y. Wu, and C. Q. Zhang, "Laplacian centrality: New centrality measure for weighted networks," *J. of Information Sciences*, vol. 194, pp. 240–253, 2012.
- [30] D. Reynolds, "Gaussian mixture models," in *MIT Lincoln Laboratory*, 244 Wood St., Lexington, MA 02140, USA, 2007.
- [31] A. Bach and V. Cabrera, "Robotic milking: Feeding strategies and economic returns," *J. of Dairy Science*, vol. 100, pp. 7720–7728, 2017.



- [32] K. Gajamannage, E. M. Bollt, M. A. Porter, and M. S. Dawkins, "Modeling the lowest-cost splitting of a herd of cows by optimizing a cost function," *J. of Nonlinear Sciences*, vol. 27, 2017.
- [33] H. Kim, S. Oh, S. Ahn, and B. Choi, "Real-time temperature monitoring to enhance estrus detection in cattle utilizing ingestible bio-sensors: Method and case studies," *J. of KIIT*, vol. 15, pp. 65–75, 2017.
- [34] A. Panisson, "Generalized random waypoint to support any number of spatial dimension," <https://github.com/panisson/pymobility>, 2012.
- [35] P. M. Fricke, P. D. Carvalho, J. O. Giordano, A. Valenza, G. J. opes, and M. C. Amundson, "Expression and detection of estrus in dairy cows: The role of new technologies," *J. of Animal*, vol. 1, pp. 134–143, 2014.
- [36] J. B. Roelofs, F. J. C. M. van Eerdenburg, N. M. Soede, and B. Kemp, "Various behavioral signs of estrous and their relationship with time of ovulation in dairy cattle," *J. of Theriogenology*, vol. 63, pp. 1366–1377, 2005.
- [37] M. L. O'Connor, "Heat detection and timing of insemination for cattle," <https://extension.psu.edu/heat-detection-and-timing-of-insemination-for-Cattle>, 2017, [Last Access Nov. 2017].
- [38] S. L. Walker, R. F. Smith, J. F. Routly, M. J. M. D. N. Jones, and H. Dobson, "Lameness, activity time-budgets, and estrus expression in dairy cattle," *J. of Dairy Science*, vol. 91, pp. 4552–4559, 2008.
- [39] A. D. Rahaman, G. Bishop-Hurley, D. Henry, and R. Rawnsley, "A comparison of autoencoder and statistical features for cattle behaviour classification," in *Neural Networks (IJCNN)*, Vancouver, BC, Canada, 2016, pp. 1232–1237.
- [40] A. J. Itle, J. M. Huzzey, D. M. Weary, and M. A. G. von Keyserlingk, "Clinical ketosis and standing behavior in transition cows," *J. of Dairy Science*, vol. 98, pp. 128–134, 2015.
- [41] M. S. Herskin, F. Skj  th, and M. B. Jensen, "Effects of hunger level and tube diameter on the feeding behavior of teat-fed dairy calves," *J. of Dairy Science*, vol. 93, pp. 2053–2059, 2010.



**Dixon Vimalajeewa** received his B.Sc in mathematics and statistics from The University of Ruhuna, Sri Lanka in 2012, and M.Sc in Computational Engineering from The Lappeenranta University of Technology, Finland in 2015.

Currently, he is a PhD student at Telecommunications Software and Systems Group (TSSG) at Waterford Institute of Technology (WIT). His research interests include data analytics, sensor-based animal phenotypes and distributed learning algorithms ([dvi-malajeewa@tssg.org](mailto:dvi-malajeewa@tssg.org))



**Sasitharan Balasubramaniam** received the bachelor's degree in electrical and electronic engineering from The University of Queensland in 1998, the master's degree in computer and communication engineering from the Queensland University of Technology in 1999, and the Ph.D. degree from The University of Queensland in 2005.

He is currently an Academy of Finland Research Fellow at the Department of Electronic and Communication Engineering, Tampere University of Technology, Finland, and an Acting Director of Research at the Telecommunication Software and Systems Group, Waterford Institute of Technology, Ireland, where he was involved in a number of Science Foundation Ireland projects. His current research interests include molecular and nanocommunications, and Internet of (bio-nano) Things. He is on the Steering Committee of the ACM NanoCom Conference which he co-founded. In 2018, he received the ACM/IEEE NanoCom Outstanding Milestone Award, and he is also the IEEE Nanotechnology Council Distinguished Lecturer. He is currently an Editor of the IEEE INTERNET OF THINGS JOURNAL, Nano Communication Networks (Elsevier), and Digital Communication Networks.



**Chamil Kulatunga** Chamil Kulatunga received his Bachelor in Electronics and Telecommunication Engineering from the University of Moratuwa (Sri Lanka) in 1999, Masters in Computer Science from the Waterford Institute of Technology (Ireland) in 2003, and PhD in Internet Engineering from the University of Aberdeen (UK) in 2009.

He was an Experienced Postdoctoral Researcher (2015-2017) under the Science Foundation Ireland funded PrecisionDairy project at the Telecommunication Software and Systems Group, Waterford Institute of Technology, Ireland. He is currently a Research Data Analyst under the Science Foundation Ireland and Origin Enterprises funded CONSUS project at the University College Dublin, Ireland.

His current research interests include distributed machine learning and data-driven agriculture in crop and dairy farming ([chamilkul@gmail.com](mailto:chamilkul@gmail.com)).



**Bernadette O'Brien** received her Bachelor (Dairy Science), Masters and Ph.D. degrees from the National University of Ireland (Cork) in 1982, 1985 and 1988, respectively.

She is currently a Principal Research Officer at Teagasc Moorepark's Animal and Grassland Research and Innovation Centre. In her current role at Teagasc, Bernadette's projects include research into innovative and sustainable systems combining automatic milking and precision grazing; increasing efficiency of traditional technologies and exploring new technology on dairy farms. This has followed from research on on-farm labour efficiency with respect to labour input profiles, alternative milking frequencies, various calf feeding frequencies, and milking efficiency in conventional parlours. Bernadette has also been successful in obtaining funding for research initiatives examining the use of precision technologies, technology platforms and computational biology to increase the economic and environmental sustainability of pasture-based production systems. She has recently established and coordinated a network of research scientists in six European countries in joint research in automatic milking. ([Bernadette.O'Brien@teagasc.ie](mailto:Bernadette.O'Brien@teagasc.ie)).



**Donagh P Berry** received his Bachelor in Agricultural Science and PhD in quantitative genetics at University College Dublin, Ireland in 2000 and 2003, respectively and a Masters in Bioinformatics and Systems Biology from University College Cork in 2012.

He is currently a senior principal investigator in quantitative geneticist at Teagasc, Ireland as well as being director of the VistaMilk Agri-Tech Research Centre. He holds professorships at three (inter)national universities. In his Teagasc capacity, he is responsible for the research on genetics in dairy cattle and is responsible for the development and implementation of genomic evaluations in dairy cattle, beef cattle and sheep in Ireland. As director of VistaMilk, he leads a team of ~200 scientists in the development and deployment of digital technologies in precision dairy production ([donagh.berry@teagasc.ie](mailto:donagh.berry@teagasc.ie)).